A STATISTICAL STUDY OF WELSH BRONZE AGE METAL ARTIFACTS

Morven N. Leese Institute of Archaeology, London

In 1977 a project was set up by the British Museum Research Laboratory and the Institute of Archaeology, London, and financed by the Science Research Council, to investigate statistical and pattern recognition techniques in relation to analyses of the composition of ancient metal artefacts. A large part of the data for this was provided by Peter Northover who has analysed all the Bronze Age metal artefacts held by Welsh Museums, for the Board of Celtic Studies of the University of Wales.

The purpose of this paper is not to present detailed results on the project, but to discuss the overall approaches that are possible, and to indicate the problems that have been encountered in trying to fulfil its aims.

The following topics will be discussed: (1) What do we hope to learn from metal analyses, in a favourable case? (2) Why provenancing is impossible, at least for the Welsh data. (3) The two approaches that can be taken, namely hypothesis testing and pattern-seeking, and why the former is preferred. (4) A description of the Welsh data, and the particular problems that arise from them. (5) An example of hypothesis testing applied to the Welsh data.

(1) WHAT DO WE HOPE TO LEARN FROM METAL ANALYSES?

Three different kinds of information may be available for an ancient metal artefact: the context in whihc it was found (where it was found, and what was found along with it); its typology (its shape, size and decoration); its chemical composition (the micro-structure of the metal and the relative amounts of the constituent elements). To see how the data on chemical composition can aid the interpretation of the other information, consider the following hypothetical example:

Example: Two groups of axes have been found, one in Ireland and one in Wales. They are typologically similar, and the context suggests that they were current during the same period. It is proposed that the two groups of axes have a common origin, their metal having been smelted from a known Irish ore; the presence of the axes in Wales indicates a physical movement of axes from one country to another, perhaps by way of trade.

In the above example it might be possible to show from the metal analyses that the two groups had similar compositions; and if the composition of metal smelted from all the possible ores were known, and one of them closely resembled the composition of the axes, then there would be some support for the theory.

If, on the other hand, the two groups could be shown to have significantly differing compositions, then alternative theories would have to be investigated, for example that the movement was not of axes but of the early metallurgists themselves, bringing with them an Irish typology but using local Welsh ores.

Thus the most ambitious aim of this kind of work is to be able to associate individual items or groups of items with particular ores, and hence to confirm or refute archaeological theories. If this is not possible (and it is shown below that it is not, at least for the Welsh data), then more limited aims have to be adopted. These involve: (i) testing whether a priori groups of items have significantly different compositions; (ii) looking for patterns in the compositional data that suggest homogeneous groups.

These two approaches are discussed below, but first the reasons for not pursuing the ore-sourcing question are given.

(2) WHY PROVENANCING IS IMPOSSIBLE FOR WELSH BRONZE AGE METAL

The 'source' of a metal artefact has to be considered as a combination of both ore and smelting technology. This is because, for a given ore, variation in the smelting technique used to extract the metal will greatly affect the final composition. Exactly what happens to each element depends on the type of smelt (ie whether the ore was roasted for example), the temperature and length of the smelt, the amount of oxygen, type of flux, etc.

It is because of these changes from raw material to finished product that the study of ancient metal is so much more difficult than the study of the composition of other materials such as flint or pottery, where provenancing has been successfully carried out in a number of cases. An example of provenancing medieval floor tiles can be found in Hansen and Sørenson (1978).

In order to identify the ore from which a metal item has been made, at least one of the following sets of information is required:

(i) Samples of the ores that might have been used, along with a model for predicting the composition of metal that would have been extracted from these ores, given the various possible smelting techniques.

(ii) Items that are known to have been made from particular ores during particular periods.

There are no excavated smelting sites from Bronze Age Wales resulting in information of type (ii). Models for the metal likely to arise from a given ore are at a very primitive stage (R.F. Tylecote, personal communication), although smelting experiments may throw some light on this problem, and ore analyses are almost completely lacking for Wales, so information of type (i) is also missing.

In conclusion there seems little chance of provenancing Welsh Bronze Age artefacts because the basic data are lacking. Nevertheless there are other questions of interest that can be illuminated by the compositional data; the two possible approaches to these are now discussed in general terms before giving examples taken from the Welsh data.

(3) THE TWO POSSIBLE APPROACHES: HYPOTHESIS TESTING AND PATTERN SEEKING

Cluster analysis is an example of pattern seeking commonly applied in archaeology; in cluster analysis we start with the compositional data and try to find groups of items that are relatively internally homogeneous in some sense but different from each other. This is done initially without reference to the other data that might be available, for example the typological data. After the clusters, as these groups are called, are found, an attempt is usually made to relate them to the archaeological data by seeing how many items in each cluster have something in common, for example, findspot.

Hypothesis testing starts from the other end, as it were; without reference to the compositional data, groups that have something in common are defined a priori. If the average compositions of these a priori groups differ more than expected on the basis of the variation between the individual items within the groups, then an explanation in terms of the grouping can be put forward.

The main problem with cluster analysis is that when there are many elements there is a correspondingly large number of potential clusters; the difficulty is knowing which really represent distinct groups. Moreover there are a large number of different algorithms for producing the clusters, and there is often no way of telling in advance which one is appropriate. Practical tests of a number of these algorithms have shown that they can give quite different answers for the same set of data. For a discussion, see Everitt (1974).

The result of this profusion of clustering algorithms and criteria for deciding which clusters to accept is that a typical cluster analysis involves a good deal of trial and error on the part of the analyst. So the final results tend to embody implicitly his background knowledge and judgement about what 'makes sense'. For this reason cluster analysis is considered a useful explanatory tool, but its results are not testable by standard statistical tests, most of which are based on the assumption that the hypothesis testing approach has been adopted.

Hypothesis testing has its own problems, one of which is that unless sensible a priori groups are chosen in the first place, worthwhile results are unlikely to emerge. So it is usually necessary to obtain the advice of an expert in the relevant background archaeology. However because an essentially conservative approach is taken, later stages involving the interpretation of the results can be performed with the assurance that the results are statistically valid.

Thus from the point of view of the statistician, it is generally preferable to adopt the hypothesis testing approach, if this is possible. Of course there are often cases when there is no basis for forming a priori groups, and then the starting point has to be from the compositional data, which is used to suggest hypotheses rather than test them.

(4) A DESCRIPTION OF THE WELSH BRONZE AGE DATA

Before illustrating the problems discussed above by means of two examples from the Welsh data, we give a brief description of the type of information available and of the problems that arise specifically from the data.

There are about 500 Welsh Bronze Age items in all. The earliest are pure copper or arsenical copper, and the rest are tin bronzes. Most of them were found in Wales or in the bordering counties of England. Each item has the following information stored in the form of a computer record:

Composition: Cu Sn Pb As Sb Ag Ni Co Au Fe Zn
Find-spot (including grid reference) Description (eg axe, spear, etc)
Function (eg chopping, cutting etc) Type (eg Migdale, Lough Ravel, etc)
Metal-working phase (EBA 1-5, MBA 6-8, LBA 9-11)

The metal-working phase is a roughly chronological sequencing of metal work within the main divisions of the Bronze Age into Early, Middle and Late.

Each of the element values is the result of averaging three spot microprobe readings. For Au and Zn the values are below the detection limit in almost all items. Since Fe is generally regarded as unreliable (because it is influenced even more than the other elements by the extractive process), and Cu can be regarded as being determined by the other elements (the total percentage from all elements is adjusted to 100%), we are in effect concerned only with Sn, Pb, As, Sb, Ag, Ni, Co.

Although the total sample size is relatively large (500), the numbers within one metal working phase can be quite small. For example in the Early Bronze Age the sample sizes are: phase 1 (5); phase 2 (14); phase 3 (33); phase 4 (26); phase 5 (15).

Thus the main problems that arise from the data are small sample sizes, and a reduction in the number of elements that can be considered because of detection limits.

(5) AN EXAMPLE OF THE HYPOTHESIS TESTING APPROACH

The metal used in phases 1 and 2 of the Early Bronze Age was arsenical copper. The presence of As in the copper at about 1% leads to harder metal than pure copper. It is generally thought that copper ores already containing the required amount of As were chosen. However the suggestion has also been made that a high As copper ore, or As itself, was added at some stage (Case (1954) and McKerrell (1978) discuss this).

In phase 3, bronze (ie copper alloyed with Sn) makes its first appearance alongside the arsenical copper. Since the metallurgical properties conferred by Sn are similar in kind but superior to those conferred by the As, it can be assumed that the early metallurgists would not continue to add both, if indeed As was added.

Under the hypothesis that in arsenical copper the As was added, one would expect that in Sn bronzes the As would be reduced to the level of just an impurity. The only valid comparison is between arsenical coppers and Sn bronzes that have the same basic impurities, otherwise any observed difference could be the result of the use of a different ore, rather than a change in alloy. The means and standard deviations of the As for two such groups are shown below:

	Mean	St. Dev.	Number	't'
Sn bronzes	1.07	0.86	10	
Phase 1-2 arsenical copper	1.44	1.64	13	0.64

The observed difference is quite large in percentage terms. However the variation of the individual items within each group is also large and the sample sizes are small. The 't' statistic is a standardised measure of the group difference that takes into account these factors.

The probability of observing a 't' value as great as 0.64 if there was no real difference is about 0.5 - in other words it is quite likely to have arisen by chance (in fact if we consider a 'one-tailed' t-test, it is only 0.25 - for a discussion of t-tests, see for example Davies (1967)). So, contrary to what might be inferred from the data purely by inspection, we cannot claim that there has been a reduction in the amount of As. Hence there is no support from the Welsh data for the hypothesis that the As was added deliberately.

This is an example of the simplest type of hypothesis; there are only two groups and one variable. The extension of this to more than one variable and more than two groups is discriminant analysis; this would be used, for instance, in testing differences in overall trace element composition.

The main problem with this approach from the statistician's point of view is deciding which hypotheses are interesting to the archaeologist.

(6) AN EXAMPLE OF THE PATTERN SEEKING APPROACH

In this example we consider the phase 3-5 Sn bronzes that contain As and Ni as impurities but not Ag. In this respect they are different from the items considered in the previous example which contained Ag but not Ni. Ignoring Co (which is too infrequent to be useful) we could use the Sb content of these items to define two subgroups, one containing Sb above the detection limit (0.02), and one in which the Sb is undetected, as Northover (1977) does. The distribution of Sb content is as follows:

% Sb	Frequency (0 — 10 — 20)
0.00 - 0.02	- - - - - - - - - - - - - - - - - - -
0.02 - 0.04	- -
0.04 - 0.06	- -
0.06 - 0.08	- -
0.08 - 0.10	
0.10 - 0.12	-
0.12 - 0.14	-
0.14 - 0.16	-
0.16 - 0.18	
0.18 - 0.20	-
0.20 - 0.22	
0.22 - 0.24	
0.24 - 0.26	
0.26 - 0.28	-
0.28 - 0.30	

The 19 items that have Sb between 0 and 0.02 are not observable, so we have no way of telling whether they represent a distinct group with zero Sb, or whether there is a continuous distribution of Sb contents which is concentrated at the low end and tails off slowly towards larger values.

Apart from inspecting the distribution, which in this case is not very informative, we can test the two suggested groups for differences in respect of some other criterion, for example geographical distribution. In this case there was no detectable difference in the geographical distribution, or in the average composition in terms of As and Ni. So there is no additional evidence, apart from the distribution above, for two separate groups.

In statistical terms the problem is one of detecting and testing the existence of separate modes, or peaks, in the data. The testing of multimodality in data is not well advanced in statistical theory, so one has to resort to ad hoc methods such as used in the above.

When many variables are considered simultaneously, as is done in cluster analysis packages, there is of course more chance of detecting separate groups than in the rather naive example above. It does however illustrate the often inconclusive nature of the pattern seeking approach.

CONCLUSION

It has been argued in the above that, from the statistical point of view, the most scientific approach to the investigation of ancient metal composition is one in which an archaeological theory can be specified in advance, and expressed in terms of what would be expected in the metal data. The degree of support for the theory can then be assessed objectively. It is also recognised, however, that sometimes an archaeological theory cannot be specified in terms of the compositional data, in which case a pattern seeking approach can be used to suggest rather than test hypotheses.

The main problems encountered in applying these approaches to the Welsh data can be summarised as: (i) the choice of appropriate theories to test; (ii) in the pattern-seeking approach, deciding in borderline cases which observed patterns are 'real'; (iii) data-related problems such as small sample sizes and high detection limits.

In spite of the above, it is considered that, even though the most obviously interesting question - ore-sourcing - is not possible at the moment, useful results are emerging from this project.

REFERENCES

CASE, H. (1954) "Studies of Irish and British early copper artefacts: second series" Man 54 (No.21).

DAVIES, O.L. (1967) "Statistical Methods in Research and Production" Oliver & Boyd, London.

EVERITT, B. (1974) "Cluster Analysis" Heinemann Educational Books, London.

HANSEN, B.A. & SØRENSON, M.A. (1978) "Provenance study of medieval, decorated floor tiles carried out by means of neutron activation analysis" Proc. Symposium on Archaeometry and Archaeological Prospecting, Bonn.

McKERRELL, H. (1978) "The use of tin-bronze in Britain and the comparative relationship with the Near East" Proc. Seminar on 'The Search for Ancient Tin' (Ed. Franklin, A.D. et al) Washington.

NORTHOVER, J.P. (1977) "Bronze in the British Bronze Age" Proc. Symposium on 'Aspects of Early Metallurgy' (Ed. Oddy, W.A.). Historical Metallurgy Society & B.M.R.L., London.